

United States Patent [19]

Horst et al.

US005710549A

[11] Patent Number: 5,710,549
 [45] Date of Patent: Jan. 20, 1998

[54] ROUTING ARBITRATION FOR SHARED RESOURCES

[75] Inventors: Robert W. Horst, Saratoga, Calif.; William J. Watson; David P. Sonnier, both of Austin, Tex.

[73] Assignee: Tandem Computers Incorporated, Cupertino, Calif.

[21] Appl. No.: 483,663

[22] Filed: Jun. 7, 1995

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 316,431, Sep. 30, 1994, abandoned.

[51] Int. Cl. 6 H04Q 1/18

[52] U.S. Cl. 340/825.5; 370/462

[58] Field of Search 340/825.5; 825.51; 370/444, 455, 462, 461, 447; 395/291, 296, 303, 650, 728, 729, 731, 732, 299, 860, 861, 862

References Cited

U.S. PATENT DOCUMENTS

3,699,524 10/1972 Norberg 340/825.5
 4,663,756 5/1987 Retterath 340/825.5

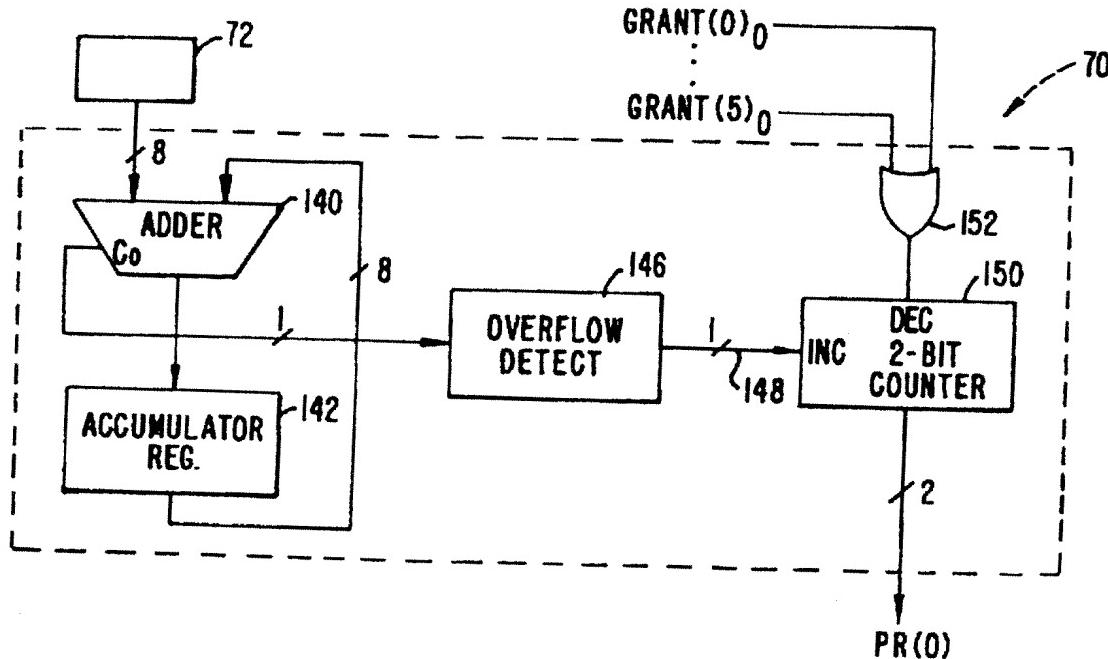
5,072,363	12/1991	Gallagher	395/725
5,210,871	5/1993	Lala et al.	395/650
5,388,097	2/1995	Baugher et al.	370/455
5,388,228	2/1995	Heath et al.	395/303
5,392,033	2/1995	Oman et al.	340/825.5
5,479,158	12/1995	Sato	340/825.5

Primary Examiner—Edwin C. Holloway, III
 Attorney, Agent, or Firm—Townsend and Townsend and Crew LLP

[57] ABSTRACT

A data communicating device, having a number of inputs whereat data is received for communication from one of a number of outputs of the device, includes apparatus for providing two levels of arbitration to select one of the inputs for data communication to an output. The first (lower) level of arbitration bases selection upon a round-robin order; the second (higher) arbitration level selects inputs based upon an indication from an input of an undue wait for access to the output over a period of time. Each input is provided a modulo-N counter and a digital counter. Each time an input contends for access to an output and loses to selection by the output to another input, the modulo-N counter is incremented by an assigned value for that input. When N is exceeded without access, the digital counter is incremented. The content of the counter operates to force the high-level arbitration.

13 Claims, 3 Drawing Sheets



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-138774

(43)公開日 平成9年(1997)5月27日

(51) Int.Cl. ⁶ G 06 F 13/362 7/50 15/173	識別記号 510	庁内整理番号 F I G 06 F 13/362 7/50 15/16	技術表示箇所 510 G K 400 T
--	-------------	---	-------------------------------

審査請求 未請求 請求項の数4 O L (全10頁)

(21)出願番号 特願平8-145734
(22)出願日 平成8年(1996)6月7日
(31)優先権主張番号 08/483663
(32)優先日 1995年6月7日
(33)優先権主張国 米国(US)

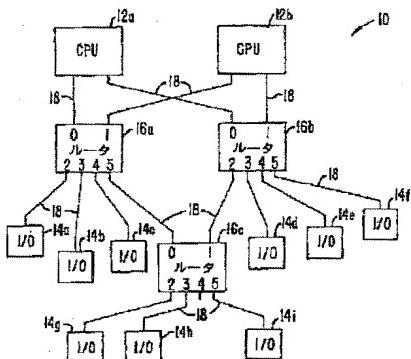
(71)出願人 391058071
タンデム コンピューターズ インコーポ
レイテッド
TANDEM COMPUTERS IN
CORPORATED
アメリカ合衆国 カリフォルニア州
95014 クーパーティノ ノース タンタ
ウ アベニュー 10435
(72)発明者 ロバート ダブリュー ホースト
アメリカ合衆国 カリフォルニア州
95070 サラトガ ラーチモント アベニ
ュー 12386
(74)代理人 弁理士 中村 稔 (外6名)
最終頁に続く

(54)【発明の名称】 共用リソースのためのルート裁定方法

(57)【要約】

【課題】 同じ出力へのデータ通信アクセスについて競合する2つ以上の入力の裁定にバイアスを与える方法及び装置を提供する。

【解決手段】 多数の出力の1つから通信するためにデータ受け取られる多数の入力を有したデータ通信装置は、出力へデータを通信するための入力の1つを選択するように2レベルの裁定を行う装置を備えている。第1の(低い)レベルの裁定は、ラウンドロビン順序に基づいて選択を行い、第2の(高い)裁定レベルは、ある時間周期にわたり出力へのアクセスが不当に待機された入力からの指示に基づいて入力を選択する。各入力には、ミュロNカウンタ及びデジタルカウンタが与えられる。入力が出力へのアクセスに競合し、それに敗れて、出力が別の入力を選択するたびに、ミュロNカウンタがその入力の指定値だけ増加される。アクセスせずにNを越えると、デジタルカウンタが増加される。カウンタの内容は、高レベル裁定を強制するように動作する。



【特許請求の範囲】

【請求項1】少なくとも一対の入力と、出力を有し、その一対の入力にメッセージデータを受け取って、出力へ通信そして出力から再送信するように動作するデータ通信装置において、出力へのアクセスに競合する一対の入力間の裁定にバイアスを与える方法が、各々の入力に指定値を与え、

一対の入力各々の値が等しいときは第1の所定のベースで出力へ通信するように一対の入力の一方を選択し、一対の入力各々の値が等しくないときは第2の所定のベースで出力へ通信するように一対の入力の一方を選択し、そして出力へ通信するためのメッセージデータを有する入力の一方の値を変更する、という段階を備えたことを特徴とする方法。

【請求項2】上記変更段階は、上記指定値により変更される各入力の累積値を発生する請求項1に記載の方法。

【請求項3】複数の入力と、少なくとも1つの出力を有し、複数の入力にデータを受け取って出力から再送信するように動作するデータ通信装置であって、出力へそしてそこから通信するためのデータを有する多数の複数の入力の中から選択を行うデータ通信装置において、複数の入力の各々に対し、

(a) 指定値を受け取りそしてそこから変更された値を発生するように接続された演算ユニットを備え、この変更された値は、上記複数の入力の1つが出力のためのデータを有そして上記複数の入力の別のものが選択されたときに上記指定値によって変更され、

(b) 上記演算ユニットに接続され、上記変更された値が所定値に等しいか又はそれを越えるときにカウントを増加するためのカウンタを備え、

上記出力は、上記複数の入力の各々からカウントを受け取って、そのカウントが第1の値であるときは第1の順序に基づき多数の入力の1つから出力へデータを通信するように多数の入力の1つを選択し、そして複数の入力のいずれかからのカウントが第1の値でないときは第2の順序に基づき多数の入力の1つを選択するためのアビタロジックを有することを特徴とするデータ通信装置。

【請求項4】上記演算ユニットは、その演算ユニットにより発生された和が桁上げを生じるときに桁上げ信号がアサートされる桁上げ出力を含む請求項6に記載の装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、一般に、多数の入力と出力を有していて通信ネットワーク内でメッセージのトラフィックをルート選択する装置に係る。より詳細には、本発明は、装置の同じ出力へのデータ通信アクセスに対して競合する装置の2つ以上の入力間の裁定にバ

イスを与える方法及びこのような方法を実施する装置に係る。

【0002】

【従来の技術】マルチ処理環境が増大するにつれて、システムの個々のシステム要素（即ち、プロセッサや周辺装置）間でデータ通信を行う機能としては、通信経路又は他の幾つかの共用リソース（例えば、マルチユーザバスシステム）への不公平なアクセスのあるシステム要素に偶発的に与えないようになると同時に、他のシステム要素へのこのようなアクセスを制限するように入念に考えねばならない。今日の多数のデータ通信ネットワーク構成は、1つの通信リンクから多数の他の通信リンクの1つへメッセージトラフィックを向け又はルート指定するために種々の装置（典型的に「ルータ」）を使用する。しかしながら、メッセージトラフィックは、同じ出力に向けて実質的に同時に装置の2つ以上の入力に受け取ることができ、2つのうちのどれを他の前に処理できるようにするか、即ちどれを最初に行かせるかの闘争が生じる。出力へのアクセスは、2つの競合する入力間にある形態の裁定を必要とする。小型のシステム（即ち、互いに通信するために少数のシステム要素しか必要としないシステム）は、入力に固定の優先順位を指定したり又は「ラウンドロビン」処理を使用したりといった古典的な裁定技術を用いることができる。前者の場合には、各ポート入力に、ハイアラーキ裁定機構を形成するための所定優先順位が与えられる。高い優先順位の入力は、その同じ出力へアクセスを求める入力であって低い優先順位が予め指定された入力以上に出力へのアクセスが与えられる。ラウンドロビン技術は、入力の間に所定の順序に基づいて優先順位を指定することを含む。メッセージトラフィックを受け取ると、優先順位が変化し、特定の出力へのアクセスが許可された最後の入力に最低の優先順位が与えられ、その次の順の入力が最高の優先順位を有し、残りの入力は、所定の順序に基づいて同様に変化する優先順位を有する。

【0003】例えば、マルチユーザVMEバスシステムのような他の共用リソース環境に同様の技術を用いて、バスへのアクセスを、接続されたコントローラ間に割り当てるともできる。

【0004】

【発明が解決しようとする課題】コンピュータシステムが複雑になり、要素（プロセッサ及び周辺ユニット）の数が増加するにつれて、ルート指定装置の入力が多数の要素からのメッセージトラフィックをルート指定するよう要求されることが例外ではなくなった。更に、ルート指定装置は、例えば、ツリー状のネットワーク構成するように時々カスケード接続にされ、ルート指定装置の入力を経てメッセージトラフィックをルート指定する必要のあるシステム要素の数が増加する。予め指定された優先順位、ラウンドロビン処理、又は他の同様の技術を

使用する場合には、ルート指定装置のある入力のサービスがシステム要素の幾つかに不公平に割り当てられ、そのルート指定装置の別の入力を使用する他の要素を犠牲にすることがある。例えば、あるルート指定装置は、1つのシステム要素のみからメッセージトラフィックを1つの入力に受け取るが、同じルート指定装置の別の入力は、多数の要素にサービスすることが要求される。上記技術のいずれかを使用すると、ルート装置の注意の多くが1つの要素に向けられ、第2の入力を用いる多数の要素の各々にはあまり注意が向けられない。従って、公知の裁定技術は、ルート指定装置のサービスの多くを、そのルート指定装置を使用する少数のシステム要素を有する入力に不公平に割り当てることになる。

【0005】

【課題を解決するための手段】本発明は、例えば、共用バス構造体のような任意の形式の共用リソース、又はここに述べるネットワークルート指定装置の出力へのアクセスを求める多数のユーザ間を裁定する方法を提供する。本発明は、多数のメッセージ受信ポート入力と、少なくとも1つの出力ポートとを有し、上記ポート入力の1つにメッセージトラフィックが受け取られて、その出力ポートヘルート指定されそしてそこから送信されるようなネットワークルート指定装置に使用される好ましい実施形態について説明する。本発明は、各ポート入力を通るメッセージトラフィックを監視して、待機中のメッセージトラフィックを有するポート入力の優先順位を上昇するという考え方方に基づいている。本発明の好ましい実施形態によれば、2レベルの裁定が実施される。最初に、低レベル優先順位機構を用いて、ポート入力へのアクセスを求める2つのポート入力間で裁定が行われ、多数の裁定周期を通じて待機していたメッセージトラフィックを有するポート入力を受け入れるために、それらの優先順位が上昇され、高い優先順位機構を瞬間に使用する裁定へと持っていく。低レベル機構は簡単なラウンドロビン優先順位を使用する。ポート出力へのアクセスをもつ最後のポート入力は、裁定の優先順位を有し、一方、次の順番のポート入力が最高の優先順位を有する。他のポート入力は、確立されたラウンドロビン順序に基づいて下降する優先順位を有する。ポート入力がポート出力へのアクセスを得るたびに優先順位が変化する。

【0006】第1レベルにおいて裁定を行うときには、各々の選択されたポート入力は、その受け取ったメッセージトラフィックを送信のために出力ヘルート指定し、優先順位がラウンドロビン機構に基づいて進んで、そのポート入力の優先順位を最低にし、そしてラウンドロビンシーケンスで次のポートを最高の優先順位をもつものとして指名する。高レベルの裁定機構は、ポート出力を通るメッセージトラフィックを監視しそしてポート出力ヘルート指定されるべく待機しているメッセージトラフ

ィックを有するポート入力の優先順位を高める。高められる量は、各ポート入力に対して予め指定されたバイアス値に一部依存する。予め指定されたバイアス値は、入力分数(I F)の形態であり、これは、ポート出力の帯域内の一部分のポート入力の割り当てとして考えられる。ポート出力により取り扱われるメッセージトラフィックの流れが監視され、待機中のメッセージトラフィックを有するポート入力の優先順位は、ポート入力が待機しているポート出力へのアクセス及び裁定に勝つことができないたびに高めることができる。そのポート入力に指定されたI Fにより時間間隔が確立される。ポート入力が出力ポートの裁定に参加する各裁定周期は、予め指定された値でモジュロNカウンタを増加する。このカウンタがオーバーフローすると、2ビットカウンタが増加される。この2ビットカウンタの非ゼロカウントは、関連ポート入力がポートへ優先順位要求を発行し、その優先順位が高められ、そして順番が変えられる(out of turn)という信号である。2つ以上のポート入力がその関連2ビットカウンタに非ゼロカウントを有する場合は、高レベル機構が、大きい方の非ゼロカウントをもつものへアクセスを許可する。2つ以上のポート入力の2ビットカウンタが非ゼロであって且つ等しい場合には、固定の優先順位機構で裁定が行われる。2ビットカウンタに非ゼロカウントを有するポート入力がアクセスを許可されるたびに、2ビットカウンタが減少される。

【0007】従って、優先順位カウンタに0以外のカウントを有するポート入を見る裁定周期が最初に裁定され、同点の場合は、固定優先順位によって行われる。各ポート入力のI F値を形成する指定のバイアス値は、モジュロNカウンタ構成を増大するのに使用される。(ここで、Nは256であり、従って、カウンタは255でロールオーバーする。) カウンタがその最大カウント(255)を越えて増加されそしてロールオーバーすると、2ビットカウンタが1だけ増加される。本発明により多数の効果が達成される。第1に、ルート指定装置のポート出力への公平なアクセスが、いかなるポート入力にも与えられ、即ちシステム要素に直結されたポート入力及びシステム要素に間接的に接続された(即ち他のポートを経て)ポート入力の両方に与えられる。第2に、本発明は、ポート出力の最大帯域内の最小保証部分をいかなるポート入力にも割り当てる。他のメッセージトラフィックを待機してデータが失われないよう確保するために、例えば、リアルタイムデータ(例えば、ビデオ)を取り扱うポート入力にポート出力へのアクセスを傾斜させるように、高い指定バイアス値を与えることができる。第3に、全ネットワークについて指定のバイアス値を与えることにより、いかなる2つのシステム要素間にも保証されたメッセージ送信待ち時間を確立することができる。これは、ネットワークの混雑によるのではないエラー状態のもとでのみ越える特定の値に低い時間切れ

値をセットできるようにする。

【0008】本発明のこれら及び他の効果は、添付図面を参照した本発明の以下の詳細な説明より当業者に容易に明らかとなろう。

【0009】

【発明の実施の形態】添付図面の図1には、簡単なマルチプロセスシステムが参考番号10で一般的に示されている。図示されたように、このマルチプロセスシステム10は、少なくとも一対の中央処理ユニット(CPU)12a、12bと、ルート指定ユニット即ちルータ16及び両方向性通信リンク18によりシステムエリアネットワーク構成に相互接続された複数の入力/出力ユニット14(14a、14b、…14i)とを備えている。システム10の種々の要素間のメッセージトラフィックは、直列送信される9ビット記号と、これら記号の同期転送のために必要な送信クロックとを含むデータパケットの形態であるのが好ましい。これらの記号は、ネットワークプロトコルの流れ制御に使用されるデータ又はコマンドを形成するようにエンコードされる。ネットワーク流れ制御は、本発明の理解又は実施に関与せず、従って、ルート指定ユニット16の幾つかの要素の説明に必要な以外は、ここでは詳細に述べない。しかしながら、各メッセージは、メッセージのソース及び行先を識別するデータを含む。行先は、メッセージが再送信されてくるところのポート出力を選択するためにルータ16により使用される。

【0010】図1の説明を続けると、各々のルータ16は、6個の両方向ポート(0、1、…5)を有し、その各々は、メッセージトラフィックが受け取られるポート入力と、メッセージトラフィックを送り出すことのできるポート出力とを有している。ルータ16aのポート2、3、4(及び0、1)の各々は、1つのシステム要素にのみサービスする(即ち、その要素からのトラフィックをルート指定する)。一方、ポート5は、このポートを経てルート指定されるメッセージトラフィックを送信できる8つのシステム要素、即ち両CPU12(ルータ16b及び16cを経て)と、6つのI/Oユニット14(ルータ16b、16cを経て)とを有する。これら8個全てのソースが、ルータ16aのポート5のポート入力を経てルート指定されるべきメッセージトラフィックを送信し、ルータ16aのポート出力、例えば、ポート0のポート出力(O)を経て送信しようとすることがある。これに対し、ルータ16aのポート1-4は、ポート0のポート出力へのアクセスに競合する必要のある单一の要素しか有していない。ポート出力への等しいアクセスが各ポート入力に許可される判定方法では、ポート2、3及び4に各々接続されたI/Oユニット14a、14b及び14c各々の方が、ルータ16aのポート5にメッセージトラフィックを送信するI/Oユニット14g、14h及び14iよりもポート0への

10 20 30 40 50

アクセスがより頻繁に与えられる。本発明は、ポート2-4よりも頻繁にポート0(又は他のポート)へのアクセスを与えるようにルータ16aのポート5をバイアスすることによりこの問題を軽減するよう作用する。

【0011】図2はルータ16aの簡単なブロック図である。ルータ16b及び16cは、特に指示のない限り、ルータ16aと実質的に同じに構成され、従って、ルータ16aについての以下の説明がルータ16b、16cにも等しく適用されることが明らかである。上記のように、ポート0、1、…5の各々は、メッセージトラフィックを送信及び/又は受信することができる。それ故、図2は、ルータ16aが、各ポート0、1、…5に対し、メッセージトラフィックを受信するためのポート入力(I)と、メッセージトラフィックを送出するためのポート出力(O)とを有するものとして示している。各ポート入力は、メッセージトラフィックの受信を処理するための関連入力ロジック30(30₀、30₁、…30₅)と、メッセージトラフィックを送出する出力ロジック32(32₀、32₁、…32₅)とを有する。到來するメッセージトラフィックは、受信ポートの入力ロジック30から、クロスバースイッチ34によりポート出力の1つヘルート指定され、クロスバースイッチは、制御・状態ロジック36(及び以下に述べる個々の出力ロジック要素32)によって一部制御される。従って、例えば、ポート0のポート入力(I)により受け取られるメッセージトラフィックは、それに関連する入力ロジック30へ送られ、そしてクロスバースイッチ34により指定の出力ロジック(例えば、出力ロジック30₀)へヘルート指定される。ポート3のポート出力O(3)は、これに接続されたデータを送信するための出力ロジック32₃を有する。

【0012】制御・状態ロジック要素は、ルータのほとんどの動作に対して同期制御を与える種々の状態マシンを含んでいる。更に、ルータ16aは、ルータの要素を同期動作するのに必要な種々のクロック信号を供給するクロックロジック40と、1つの例外を除いてここでは本発明に関連しない幾つかの自己チェック動作を実行するための自己チェック回路42とを備えている。ルータ16aには、これをメンテナンス処理システム(図示せず)へ通信接続するためにオンラインアクセスポート(OLAP)46が設けられる。このOLAP46は、以下に述べるように、ルータが、例えば、各ポート入力の指定のバイアス値のような種々の動作情報を受信できるようにするインターフェイスをメンテナンス処理システムに与える。OLAP46は、IEEE規格1149.1に合致するよう構成されたシリアルバス48に接続される。従って、始動時又は動作進行中に情報がルータ16aに与えられる。当業者に明らかなように、IEEE規格1149.1は、IEEE規格1149.1-1990年版、1990年5月21日、SH1314

4、インスティテュート・オブ・エレクトリカル・アンド・エレクトロニック・エンジニア、345イースト、47番ストリート、ニューヨーク、ニューヨーク州、10017をベースとするものである。更に詳細な情報は、この規格を参照されたい。

【0013】図3は、ポート入力I(0)の入力ロジック30₀を示すブロック図である。他のポート入力I(1)、…、I(5)の入力ロジック30₁～30₅も実質的に同じ構造であり、特に指示のない限り、入力ロジック30₀の説明を、入力ロジック30₁～30₅の説明と考えられたい。図3に示すように、入力ロジック30₀は入力レジスタ50を備え、これは、到来するメッセージトラフィックを受信してバッファし、入力の先入れ先出しバッファ待ち行列(FIFO)52へ転送するように動作する。FIFO52は、送信エンティティにおいて発信されてレジスタ50及びFIFO52へのデータをクロックするのに使用されるクロック信号(図示せず)と、FIFO52から記号を引っ張るのに使用される(ローカル)クロックとの間の同期を与えるように動作する。入力FIFO52からの情報は、9-8(ビット)コンバータ54へ送られ、これは、各々の9ビット記号をそのエンコード形態からバイト形態へ変換する。更に、入力FIFO52からの出力は、コマンドデコード要素56及びプロトコル・パケットチェックユニット58へ接続される。コマンドデコードユニット56は、各記号を検査して、それが流れ制御コマンドであるかどうか、ひいては、ルータが作用を与えねばならないコマンド、又はルータにより作用を受ける必要のあるデータ(適切なポート出力ヘルート指定するのではない)であるかどうか判断する。プロトコル・パケットチェックユニット58は、パケットが必要な転送プロトコルを満足するよう確保すると共に、パケットの結果のチェック和をチェックして、パケットがルータ16aへ適切に送信されるよう確保するように動作する。もしそうでなければ、プロトコル・パケットチェックユニット58は、パケットの終わりに、そのパケットをおそらく誤りであると識別する記号を附加する。

【0014】9-8ビットコンバータ54を通過した到来メッセージトラフィックは、FIFO制御器64によって制御される弾力性FIFO62に受け取られて一時的に記憶される。FIFO62は、到来メッセージパケットの行先IDを検査できるようにし且つポート出力がクロスバースイッチ34を操作してメッセージトラフィックをルート指定する時間を許すに充分な一時的記憶容量を備えている。又、FIFO62は、受信ポート入力が待機しなければならない場合には到来メッセージトラフィックの送信を停止するに充分な時間を許すに足る記憶容量を備えていなければならない。しかしながら、適切なポート出力の選択は、到来メッセージパケットに含まれた行先アドレスによって左右される。この決定は、

到来メッセージパケットに含まれた行先アドレスを受け取るポート出力選択ロジック66により行われる。この行先アドレスから、ポート出力選択ロジック66は、指定のポート出力を識別し、6本の要求ラインR(O)mの1つにその要求されたポート出力を識別する要求信号をアサートする。但し、n=0、1、…、5である。説明を続ける前に、表示法について述べる。上記のように、出力ポート選択ロジック66は、6本の要求ラインR(O)₀、R(O)₁、…、R(O)₅の1つにおいて各自搬送される6個の出力信号を発生する。要求信号ラインの形態は、R(n)_mであり、但し、n(n=0、1、…、5)は、信号ラインのドライブソースを識別し、そしてm(m=1、2、…、5)は、搬送される信号の行先を識別する。従って、ポート出力選択ロジック66は、6本の要求ラインR(O)₀、R(O)₁、…、R(O)₅を駆動し、その各自は、それが搬送する信号を出力ロジック32₀、32₁、…、32₅へと各自接続する。同様に、各ポート出力の出力ロジック32は、受信した要求信号に応答して、6本の信号ラインGRANT(n)_mの1つにGRANT信号をアサートすることによりアクセスを許可する。この場合も、nは信号ラインを駆動する出力ロジックを識別し、そしてmはその駆動信号を受信する入力ロジックを識別する。特に指示のない限り、この説明全体を通じてこの表示法を使用する。

【0015】図3の説明を続けると、ポート入力I(0)により受け取られ、例えば、ポート出力O(3)を識別する行先アドレスをもつ到来メッセージは、要求信号ラインR(O)₃に要求をアサートして、ポート出力O(3)(より正確には、関連出力ロジック32₃)に、該ポート出力に向けられたメッセージトラフィックがポート入力I(0)に待機していることを知らせる。この要求信号を受け取るポート出力は、次いで、そのアクセスを許可することを表す許可信号を許可信号ラインGRANT(3)に発生することで応答する。要求された出力ロジック32がアクセスを許可すると(以下に詳細に述べる)、クロスバースイッチ34を通る指定のルートが形成され、メッセージパケットは、弾力性FIFO62から要求された出力ロジックヘルート指定される。又、入力ロジックは、バイアスレジスタ72の内容を受け取るバイアスロジック70も備えている。バイアスレジスタ72は、上記のように、その関連ポート入力I(0)に対する指定のバイアス値であって、ポート出力の帯域域についてのそのポート入力の部分を表す指定のバイアス値を受け取る。バイアスレジスタ72の内容から、バイアスロジック70は、ポート入力I(0)(待機中のメッセージトラフィックをもつ)が参加して負け、その優先順位を実際に加速する優先順位要求を発生するところの裁定を監視する。この優先順位要求は、6つ全部のポート出力の入力ロジック32に接続された

2ビットバスPR(0)により所望のポート出力の出力ロジック32へ通信される。バイアスロジックは、6つのポート出力の出力ロジック32から、GRANT信号ラインGRANT(n)により搬送される許可信号を受け取る。

【0016】2つ以上のポート入力I(0)…I(5)が、同じポート出力(例えば、O(3))を識別する行先アドレスをもつメッセージトラフィックをほぼ同時に受信し始める場合には、どのポート入力を最初に処理しそしてどれを待機させるかについてある決定を行わねばならず、即ち所望のポート出力へのアクセスを裁定して、どのポート入力を最初に行かせそしてどれを待機させるかを決定しなければならない。本発明によれば、裁定は、2つのレベルで行われる。最初に、低レベル裁定が使用され、競合するポート入力が単純なラウンドロビンプロセスによって選択される(が、例えば、固定の優先順位を指定するような他の裁定構成も使用できることが明らかであろう)。多数の裁定及び要求を通じて待機したメッセージトラフィックを有するポート入力が順番を変える優先順位要求を発行することにより高レベル優先順位機構に入る。ラウンドロビン裁定プロセスは、各ポート出力O(0)、O(1)、…O(5)により、ポート出力ヘルート指定されるべき待機中のメッセージトラフィックを有するポート入力から受け取った要求信号R(n)に応答して実施される。ポート入力信号がその関連優先順位要求をアサートすることにより順番を変えられるべきであるときは、高レベル裁定機構が強制動作される。明らかなように、ポート入力が、ポート出力ヘルート指定されるべく待機しているメッセージトラフィックを有するときには、そのポート出力に対して参加する裁定を監視する。待機が続くと、各ポート入力に与えられて入力ロジック30のバイアスレジスタ72(図3)に維持された上記の入力分数(IF)からバイアスロジック70により優先順位要求が発生される(以下に述べるように)。

【0017】各ポート入力(I(0)、I(1)…I(5))からの2ビット優先順位要求は、優先順位要求バス(PR₀、PR₁…PR₅)によってポート出力(O(0)、O(1)…O(5))へ接続される。多数のポート入力が1つのポート出力に対して待機中のメッセージトラフィックを有し、それ故、そのポート出力へのアクセスを張り合っており、そしてそれらの各々の優先順位要求がゼロである場合には、裁定が行われ、競合するポート入力の1つがラウンドロビンプロセスを用いて選択される。一方、張り合っているポート入力の1つが非ゼロの優先順位要求を発行する場合には、そのポート入力が高い優先順位を有するものとして処理され、次の裁定周期中にアクセスの順番変更が許可される。2つ以上のポート入力が順番の変更を要求している場合には、固定の優先順位ベースで非ゼロの基準要求を

有するポート入力間でポート出力により裁定が行われる。優先順位要求をいかに発生して使用するかを説明する前に、先ず初めに、ポート出力(O(0)、O(1)…O(5))のアーキテクチャーを理解するのが有用であろう。図4は、ポート出力O(3)の出力ロジック32のアーキテクチャーを簡単な形態で示すものである。他のポート出力O(0)～O(2)及びO(4)～O(5)に対する出力ロジック32は、実質的に同じ構造である。図4に示したように、クロスバースイッチ34の出力は、出力ロジック32のマルチブレクサ(MUX)80によって受け取られ、これは、クロスバースイッチ34からのデータ及びコマンド信号発生器82の出力を選択するように動作する。使用されるネットワークプロトコルに基づき、制御・状態ロジック36(図2)の指令及び制御のもとで、コマンド記号を周期的に挿入し、送信することが必要となる。MUX80により行われる選択は、出力レジスタ84へ接続され、そしてそこからポート出力O(3)を経、そしてポート3が接続されたネットワーキング18(図1)を経てI/Oユニット14へ送られる。

【0018】裁定は、各ポート出力において、アビタ86により行われる。アビタ86は、各ポート入力I(0)、I(1)…I(5)から、対応するポート出力選択ロジック66(図3)からの要求信号ラインR(n)を受け取る。3つ以上の要求信号が同時にアサートされた場合には、アビタロジック86は、要求を発しているポート入力の優先順位要求信号をチェックする。全てが非ゼロである場合に、アビタロジック86は、ラウンドロビン機構の優先順位に基づいて要求を裁定する。しかしながら、競合するポート入力の1つが順番の変更を要求していることがその関連優先順位要求バス(例えば、ポート2の入力ロジック30に対するPR(2))上の非ゼロ値によって指示される場合には、アビタロジック86がそのポート入力へのアクセスを許可する。2つ以上のポート入力がその優先順位要求をアサートする場合には、アビタロジック86が高レベル優先順位機構に基づいてアクセスを裁定する。2ビットの優先順位要求が等しい場合には、ルート選択が固定優先順位に基づいて行われ、最も高い予め指定された優先順位をもつポート入力にアクセスが許可される。1つの2ビット優先順位バスの値が他のものよりも数値的に大きい場合には、その大きな優先順位要求をアサートするポート入力が次に選択される。

【0019】裁定が行われると、アビタロジック86は、6本の信号ライン(各々対応するポート入力の入力ロジック30に接続される)の1つを経て裁定に勝ったポート入力I(0)…I(5)へGRANT信号を発生する。更に、アビタ86は、クロスバースイッチ34へ選択信号(SEL)を発生し、選択された入力ロジック30が出力ロジックヘルート指定するようにさせ

る。図5は、ポート入力I(0)に対する入力ロジック32。のバイアスロジック70を詳細に示しており、8ビット加算器140と8ビット累積レジスタ142との組合せが含まれて、実際に、自走モジュロー255カウンタを形成する。加算器140は、対応するポート入力（ここではポート入力I(0)）に指定されてバイアスレジスタ72により維持されたバイアス値を受け取り、そのバイアス値を累積レジスタ142の内容に加算する。加算器140により形成された和は累積レジスタ142へ返送され、その内容をIF値だけ増加する。又、累積レジスタの内容は、入力ロジック30。が参加する各裁定周期ごとにIF値だけ増加される。累積レジスタ142の内容が、加算器140の巾を越える（即ち、255より大きい）点まで増加されると、加算器140の桁上げ(Co)出力にオーバーフロー信号が発生する。加算器140からのオーバーフロー信号は、オーバーフロー検出回路146へ送られ、出力(OV)に応答オーバーフロー信号をアサートし、これは、次いで、信号ライン148によって2ビットカウンタ150の増加(INC)入力へ接続される。従って、カウンタ140の検出されたオーバーフローは、2ビットカウンタ150を増加するように働く。2ビットカウンタ150の内容は、2ビット優先順位要求値を形成し、これは、2ビット優先順位要求バスPR(0)によって入力ロジック30。から6つのポート出力O(0)、O(1)…O(5)の出力ロジック32へ搬送される。

【0020】説明を続ける前に、ポート入力I(0)に割当てられる入力分数(IF)は、分子を構成するバイアスレジスタ72に含まれたバイアス値と、実際には分母である累積レジスタ142のサイズとで形成されることに注意するのが有用である。従って、ポート入力I(0)の場合、レジスタ72に保持されたバイアス値が64（図6について以下に説明する例で使用する値）である場合、ポート入力I(0)の入力分数は、 $64/256$ 、即ち $1/4$ となる。2ビットカウンタ150は、6入力のオアゲート152の出力を受け取る減少(DEC)入力を含む。GRANT信号は、ポート出力の各々からオアゲート152へ送られ、入力としてそこに付与される。関連するポート入力（即ちポート入力I(0)）が、2ビットカウンタ150に非ゼロ値を含む状態でポート出力へのアクセスの裁定に参加しそして裁定に勝った場合には、それによりポート出力から生じるGRANT信号が2ビットカウンタ150を減少させる。2ビットカウンタ150は、アンダーフローしないように、即ちカウンタの内容がゼロ値であるときに、DEC入力がオアゲート152の出力を無視しないよう設計される。

【0021】2ビット優先順位要求バスPR(0)は、他のポート入力I(1)、…、I(5)からのバスと共に、6つの優先順位バスPR(n)（n=0、1、…

…5）を形成し、これらは、ポート入力からの優先順位要求を各ポート出力のアビタロジックユニット86（図4）へ接続する。又、ここに説明するように、アビタロジックユニットは、36本の要求ラインR(n)も受け取り、その6つの各々は、6つのポート入力の各々からのものであって、どのポート入力がアクセスに対して張り合っているかをポート出力に識別する要求信号を搬送する。アビタロジック86は、一般的な従来設計の組合せロジック回路（又はプログラム可能なロジックアレー（PLA）エлемент）であり、競合するポートの優先順位バスPR(0)、…、PR(5)によって搬送される優先順位要求の状態から、どれがアクセスを受けるべきかを決定し、そしてそのアクセスを、上記のようにクロスバースイッチ34へ付与されるSEL信号によりルート指定するように構成される。いずれの2ビットカウンタにもカウントがない場合には、アビタロジックユニットは、ラウンドロビンプロセスに基づいて動作し、そのプロセスによりどれが最後にアクセスを許可されそしてどれが次の順番であるかに基づいて競合ポートの1つを選択する。一方、1つ以上の競合ポートが優先順位要求信号をアサートした場合には、最も高い優先順位要求を有するもの（即ち2ビットカウンタ150が最も高いカウントを有するポート入力）にアクセスが許可される。2ビットカウンタ150が同じカウントを含んでいる2つ以上のポート入力の優先順位要求間が同点である場合には、アビタは、固定優先順位機構を課し、選択されたポート入力にGRANT信号を発生する。

【0022】低レベルラウンドロビン裁定は、通常のメッセージトラフィックに対して使用され、高レベル裁定は、ポートが裁定に参加して不成功であった回数と、指定されたIF値とに基づいて強制的に作用される。高優先順位機構は、ポートの2ビットカウンタ150が非ゼロカウントを含むときに入る。図6を参照して本発明の動作を説明する。図6は、裁定周期TからT+8及びそれ以上に対しポート0、1及び2（同じポート出力へのアクセスを求める）の裁定を示す。最も左のカラムは、各裁定周期を識別し、他のカラムは、各裁定周期中のレジスタ142の内容及び2ビットカウンタの内容（かっこで示す）を表している。裁定周期中に裁定に勝つポートは、ダークの累積値で示されている。ポート0、1及び2の各々に割り当てられたIF値は、各カラムの上部にかっこで示されている。（ここで、出力ポートの帯域巾は、加算器140及び累積レジスタ142により形成された「カウンタ」のオーバーフロー値で示される。当業者に明らかなように、ポート出力の帯域巾の分割をいかに微細に又は粗くするかと、メッセージトラフィックに対する最大待機とに基づいて、他の値も使用することができる。）

図6は、ポート0、1及び2に対する到来メッセージト

ラフィックのみが特定のポート出力（例えば、ポート出力O(4)）に対して張り合っていると仮定する。明瞭化のため、他のポートは参加しないと仮定し、従って、図示されていない。更に、メッセージトラフィックは各ポートにスタックされ、即ちあるポートに対する到来メッセージトラフィックが裁定されて、再送信のためにポート出力O(4)にルート指定されたときに、別の到来メッセージが存在するものと仮定する。

【0023】図6を参照すれば、最初に、第1の裁定周期Tの前の時間(T-t)に、ポート0、1及び2のレジスタ142の内容は空である。従って、3つ全ての入力ポート0、1及び2がポート4に向けられたメッセージトラフィックを有すると仮定すれば、ラウンドロビン機構において第1のものである（いずれのカウンタ150にもカウントはない）ポート0が裁定周期Tの裁定に勝つ。この裁定周期の終わりに、ポート0、1及び2の各レジスタ142は、それらの指定バイアス値だけ増加され、従って、次の裁定周期T+1については、バイアス値が図示されたようになる。オーバーフローはなくそして関連2ビットカウンタ150は空のままであるから、裁定周期T+1のラウンドロビン裁定は、ポート4へのアクセスに対しラインの次のポート、即ちポート入力1（図6に太字で示す）を選択する。レジスタ142は、再び増加される。ここで、ポート2のレジスタ142はオーバーフローを経験してゼロに戻り、それに関連する2ビットカウンタ150は、「1」に増加される。従って、次に続く裁定周期T+2の間に、アービタロジック86'（ポート4のポート出力O(4)の）は、ポート2が2ビットカウンタ150にカウントを有するが、他のものは有していないことに注目し、それ故、アービタ周期T+2は、ポート2の選択を生じさせる。この裁定周期の終わりに、2ビットカウンタ150が1だけ減少され、全てのレジスタ142は、指定のバイアス値だけ再び増加される。

【0024】裁定周期T+3は、カウンタ150にカウントがないのを見て、再び、ラウンドロビン機構に基づいて、ラインの次のポート、即ちポート2にアクセスを許可する。この場合も、レジスタ142は、増加される。裁定周期T+4は、全てのレジスタ142がゼロにロールオーバーしてオーバーフローを生じ、全てのカウンタ150が「1」のカウントを含むのを同時に見る。カウンタ150のカウントは全て等しく（そして非ゼロであり）、従って、例えば、最初にポート0を見、次いで、ポート1を見、等々といつてポート4までラインを下るような固定の優先順位機構に依存することにより、同点状態を突破される。（明らかに、ポート5と他のポートとの間のように、他のポートが常に勝つ。）従って、この場合、ポート0が裁定に勝つ。レジスタ142は、対応する指定のバイアス値で再び増加され、一方、ポート0の2ビットカウンタは、当該ポート出力からD

ECREMENT信号ラインに信号を発生することにより1だけ減少される。裁定周期T+5は、ポート1及び2がそれらのカウンタ150に1のカウントを依然含んでいるのを見、他のポートは同点である。この同点状態も、固定の優先順位機構を使用することにより突破され、このときには、ポート2よりも高い固定優先順位を有するポート1が選択され、そしてそのカウンタ150が1だけ減少される。

【0025】裁定周期T+6は、ポート2のカウンタが「2」のカウントに増加されるのを見る。これは、そのときカウンタ150にカウントを有する唯一のものであるから、ポート4へのアクセスが再び得られ（同時に、低レベルのラウンドロビン機構の次のものであるが）、そしてポート2のカウンタ150が減少される。裁定周期T+7は、ポート2がそのカウンタ150に非ゼロカウントを有している唯一のものとして見、それ故、ポート4へのアクセスが再び選択され、そのカウンタが減少される。この裁定周期の終わりに、ポート0、1、2のレジスタ142が増加されると、全てがゼロ値へとロールオーバーそして全てがそのカウンタ150に「1」のカウントを有する。裁定周期T+8及びそれに続く周期は、ここで、裁定周期T+4、…T+7を繰り返す。図6を検討することにより明らかなように、このパターンは、ポート2が回数 $(128 / (64 + 64 + 128)) = 128 / 256$ の半分だけ裁定に勝つことを示している。一方、ポート0及び1の各々は、回数 $(64 / 256)$ の1/4だけ裁定に勝つ。従って、この機構は、指定バイアス値と2ⁿの比に基づいて帯域幅を割り当てるのに使用され、ここで、nは、バイアスされた裁定カウンタ142の巾（ここでは、8ビット）である。しかしながら、異なる巾のカウンタを使用して、カウンタ142を実施し、帯域幅を割り当てるための比に更に大きな分解能を得ることもできる。更に、6個のポートを有するルータの場合には、2ビットカウンタで充分であるが、6ポートより多くのポートを有するルータは2ビットより大きなカウンタ150を必要とする。

【0026】ラウンドロビン裁定を用いるのではなく、低レベル機構を実施する他の方法もあり、ここに開示した高レベル機構と共に、固定優先順位を使用するか又はメッセージ自体の情報を使用して裁定を行うことができる。更に、高レベル機構において同点状態を打破するために使用される固定優先順位機構は、種々の形態の組合せロジック（例えば、ゲート、プログラム可能なロジックアレー、ロックアップテーブル、等）で実施される他の予め定められた優先順位へ変更することができる。

【図面の簡単な説明】

【図1】一対の中央処理ユニット（CPU）を備え、これらは、互いに接続されると共に、本発明によるルート指定装置を用いてメッセージトラフィックを通信するためのシステムエリアネットワーク（SAN）により複数

の入力／出力（I/O）ユニットにも接続されるようなマルチプロセッサシステムの簡単なブロック図である。

【図2】図1のシステムエリアネットワークに使用されるルート指定装置の簡単なブロック図であって、メッセージトラフィックを受け取りそして再送信する多数の個別の入力及びポート出力を有する構造を示す図である。

【図3】図2に示されたルート指定装置の1つのポート入力に関連した入力ロジックを示す簡単なブロック図である。

【図4】図1及び2のルート指定装置の1つのポート出力に関連した出力ロジックを示す簡単なブロック図である。

【図5】待機中メッセージトラフィックを有する図2のポート入力の優先順位を上昇するための優先順位要求を発生するのに使用されるロジックを示すブロック図である。

【図6】メッセージトラフィックを有するポート入力が出力へのアクセスに対して裁定される多数の裁定周期を示す図である。

【符号の説明】

10 構成要素

* 12 a, 12 b CPU

14 入力／出力ユニット

16 ルート指定装置（ルータ）

18 両方向性通信リンク

30 入力ロジック

32 出力ロジック

34 クロスバースイッチ

36 状態ロジック

40 クロックロジック

10 42 自己チェック回路

46 オンラインアクセスポート（OLAP）

50 入力レジスタ

52 先入れ先出しバッファ待ち行列（FIFO）

56 コマンドデコード要素

58 プロトコル・パケットチェックユニット

62 弹力性FIFO

66 出力ポート選択ロジック

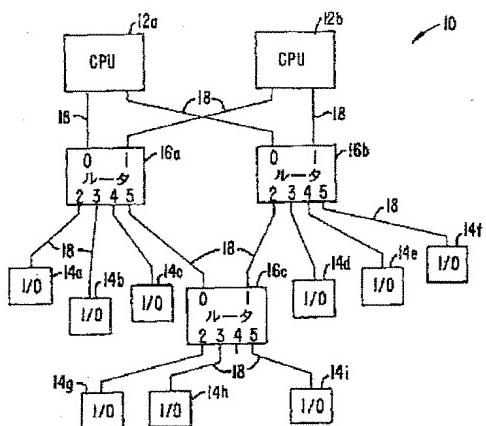
70 バイアスロジック

80 マルチプレクサ

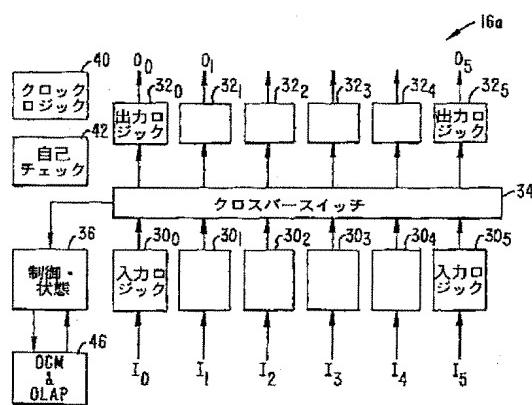
20 82 コマンド信号発生器

*

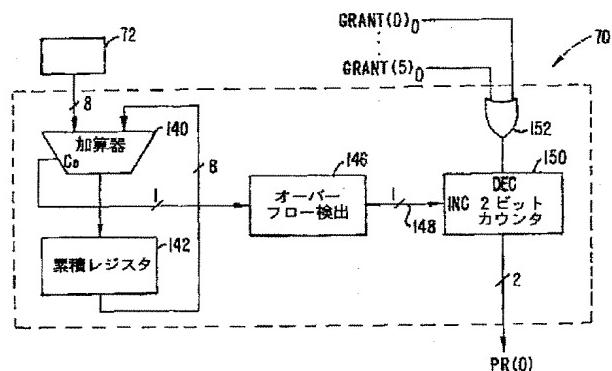
【図1】



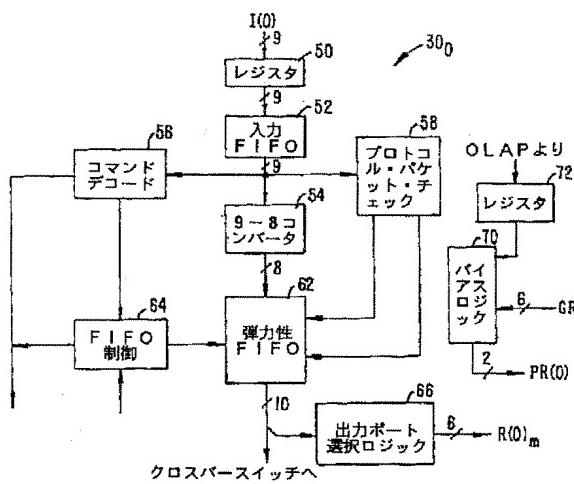
【図2】



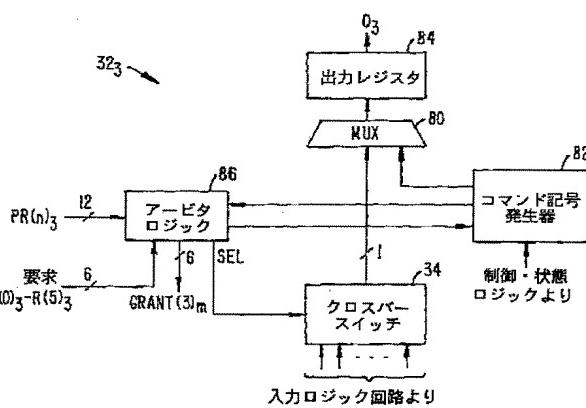
【図5】



【図3】



【図4】



【図6】

裁定周期	(IF=64/256) ポート 入力 I(0)	(IF=64/256) ポート 入力 I(1)	(IF=128/256) ポート 入力 I(2)
	T-1	0 [0]	0 [0]
T	0 [0]	0 [0]	0 [0]
T+1	64 [0]	64 [0]	128 [0]
T+2	128 [0]	128 [0]	0 [1]
T+3	192 [0]	192 [0]	128 [0]
T+4	0 [1]	0 [1]	0 [1]
T+5	64 [0]	64 [1]	128 [1]
T+6	128 [0]	128 [0]	0 [2]
T+7	192 [0]	192 [0]	128 [1]
T+8			

裁定周期 T + 4 から T + 7 を繰り返す

フロントページの続き

(72)発明者 ウィリアム ジョエル ワトソン
アメリカ合衆国 テキサス州 78756 オ
ースチン ウルリク アベニュー 1501

(72)発明者 ディヴィッド ポール ソニア
アメリカ合衆国 テキサス州 78750 オ
ースチン イメージ コーブ 7804